JIPB Journal of Integrative Plant Biology

Transcriptome analysis of near-isogenic lines provides molecular insights into starch biosynthesis in maize kernel

Yingni Xiao[1], Shawn Thatcher[2], Min Wang[1,3], Tingting Wang[1], Mary Beatty[4], Gina Zastrow-Hayes[4], Lin Li[5], Jiansheng Li[1], Bailin Li[2] and Xiaohong Yang[1]*

[1]National Maize Improvement Center of China, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China, [2]DuPont Pioneer, 200 Powder Mill Road, Wilmington, DE 19880, USA, [3]College of Agronomy, Northwest Agricultural and Forest University, Yang Ling 712100, China, [4]DuPont Pioneer, Johnston, Iowa 50131, USA, [5]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, Minnesota 55108, USA. *Correspondence: yxiaohong@cau.edu.cn

Research Article

Abstract Starch is the major component in maize kernels, providing a stable carbohydrate source for humans and livestock as well as raw material for the biofuel industry. Increasing maize kernel starch content will help meet industry demands and has the potential to increase overall yields. We developed a pair of maize near-isogenic lines (NILs) with different alleles for a starch quantitative trait locus on chromosome 3 (*qHS3*), resulting in different kernel starch content. To investigate the candidate genes for *qHS3* and elucidate their effects on starch metabolism, RNA-Seq was performed for the developing kernels of the NILs at 14 and 21 d after pollination (DAP). Analysis of genomic and transcriptomic data identified 76 genes with nonsynonymous single nucleotide polymorphisms and 384 differentially expressed genes (DEGs) in the introgressed fragment, including a hexokinase gene, *ZmHXK3a*, which catalyzes the conversion of glucose to glucose-6-phosphate and may play a key role in starch metabolism. The expression pattern of all DEGs in starch metabolism shows that altered expression of the candidate genes for *qHS3* promoted starch synthesis, with positive consequences for kernel starch content. These results expand the current understanding of starch biosynthesis and accumulation in maize kernels and provide potential candidate genes to increase starch content.

Keywords: Coexpression network; maize; quantitative trait loci; RNA-Seq; starch content
Citation: Xiao Y, Thatcher S, Wang M, Wang T, Beatty M, Zastrow-Hayes G, Li L, Li J, Li B, Yang X (2016) Transcriptome analysis of near-isogenic lines provides molecular insights into starch biosynthesis in maize kernel. **J Integr Plant Biol** XX:XX–XX doi: 10.1111/jipb.12455
Edited by: Zhongchi Liu, University of Maryland, USA
Received Aug. 4, 2015; **Accepted** Dec. 14, 2015
Available online on Dec. 17, 2015 at www.wileyonlinelibrary.com/journal/jipb

INTRODUCTION

Maize (*Zea mays*) is the world's largest crop and serves as an important source of human food, animal feed and biofuel. In maize kernels, starch is the most abundant component, comprising approximately 70% of dry kernel weight. In addition to the direct consumption of food or feed, maize starch also plays a vital role in industrial applications such as bio-ethanol production and the miller industry (Ellis et al. 1998). Thus, enhancing starch content may not only lead to higher maize kernel yields but could also increase the portion used in industrial applications.

Starch in maize kernels is composed of two types of glucan polymers: amylopectin and amylose. In many cases, amylopectin accounts for approximately 75% of the total starch content in normal maize (Myers et al. 2000) and is composed of branched glucose chains that form insoluble, semi-crystalline granules. Amylose is composed of linear chains of glucose that adopt a helical configuration within the granules (Martin and Smith 1995). Although the glucan composition of starch is rather simple, starch metabolism is relatively complex in maize (Figure S1). Starch biosynthesis initiates with the degradation of sucrose by sucrose synthase, then at least four classes of enzymes contribute to amylopectin or amylose biosynthesis: ADP-glucose pyrophosphorylase (AGPase), starch synthase, starch branching enzyme, and starch debranching enzyme (Jeon et al. 2010; Zeeman et al. 2010). In addition to debranching enzymes, a series of α- and β-amylases also degrade starch in the endosperm during cereal germination. In maize, dozens of key genes play major roles in starch biosynthesis, including *Shrunken1* (*Sh1*), *Shrunken2* (*Sh2*), *Brittle1* (*Bt1*), *Brittle2* (*Bt2*), *Waxy1* (*Wx1*), *Dull1* (*Du1*), *Amylose extender1* (*Ae1*), *Sugary1* (*Su1*), *Sugary2* (*Su2*), *SBEIa*, *SBEIIa* and *Zpu1* (Hennen-Bierwagen et al. 2013). Although the general process of starch metabolism in maize is well understood, its regulation remains unclear.

To fully understand starch biosynthesis and accumulation in maize kernels, substantial effort has been undertaken to identify quantitative trait loci (QTL) for kernel starch content (Goldman et al. 1993; Séne et al. 2000; Wassom et al. 2008; Zhang et al. 2008; Yang et al. 2013a). However, so far all QTLs for starch content have been mapped to large regions, preventing the identification of causative genes. Coexpression analysis has also proven to be a powerful tool to systemically identify genes that regulate specific metabolic pathways at the transcriptional

level (Yonekura-Sakakibara et al. 2008; Fu and Xue 2010). This analysis assumes that genes with similar expression patterns are more likely to be functionally associated. In maize, much has been learned about the systemic regulation of the biosynthesis and accumulation of maize kernel components, including starch, at various developmental stages (Lai et al. 2004; Lu et al. 2013; Chen et al. 2014). Nevertheless, the exact molecular processes that activate starch metabolism in maize kernels largely remain to be elucidated.

Maize kernel development is usually characterized by three partially overlapping stages: the lag phase during 0–15 d after pollination (DAP), the kernel-filling period from approximately 15 DAP to 40–50 DAP, and the desiccation–maturation stage after approximately 40 DAP (Prioul et al. 2008). Maize starch begins to accumulate during the kernel-filling period, with the greatest increase in kernel size occurring around 20 DAP (Prioul et al. 2008). The parallel transcription of genes related to starch metabolism is most active during the lag phase but begins to decrease around 20 DAP, midway in the filling stage (Méchin et al. 2007; Prioul et al. 2008). Previously, a major QTL for starch content, *qHS3*, explaining 26% of starch content variation, was identified in a B73 × DHLoPro1 $F_{2:3}$ population (unpublished results). This QTL was also detected in three additional linkage populations (Goldman et al. 1993; Wassom et al. 2008; Yang et al. 2013a), indicating the importance and stability of *qHS3*. To facilitate the cloning of *qHS3* and give insight into the regulatory network for starch-related genes caused by the candidate genes of *qHS3*, we produced a pair of near-isogenic lines (NILs) that varied in kernel starch content by approximately 3%. The developing kernels of the NILs at 14 and 21 DAP were subjected to RNA sequencing (RNA-Seq) to identify differentially expressed genes in the interval of *qHS3* on chromosome 3 and to elucidate the transcriptional effects of the candidate genes in this QTL on starch metabolism pathways.

RESULTS

Characterization of NILs

A major QTL for starch content on chromosome 3, *qHS3*, accounting for 26% of starch content variation in an $F_{2:3}$ population derived from B73 with 68% of the kernel starch content and DHLoPro1 with 75% of the kernel starch content, was mapped between two SNP markers PZA1 and PZA2, which have physical positions of 113 Mb and 175 Mb, respectively, according to B73 reference genome sequences in Version 5b.60 (unpublished results). We then developed a pair of BC_5F_2 NILs, NIL_B, which carries the high starch allele of *qHS3* from DHLoPro1 in the B73 background, and its matching line, NIL_A, carrying the homologous segment from B73. To define the fragments introgressed from DHLoPro1 in the NILs, the two NILs, together with their parents, DHLoPro1 and B73, were genotyped using MaizeSNP3K array containing 3,072 single nucleotide polymorphisms (SNPs). A total of 1,323 high-quality SNPs are polymorphic between the two parents, with the number of SNPs distributed on each chromosome ranging from 66 (chromosome 10) to 240 (chromosome 8). NIL_A carries a approximately 26 Mb fragment from DHLoPro1 while NIL_B carries a 175 Mb fragment from DHLoPro1 on chromosome 3, with a 13 MB overlap (Figure 1A). The two NILs differ from 15 Mb to 193 Mb on chromosome 3, with 4.1% (54/1323) of SNPs polymorphic between NIL_A and NIL_B, containing the *qHS3*.

As expected, kernel starch content is significantly higher in NIL_B (average 72.1%) than NIL_A (average 69.3%) (Figure 1B). No significant difference in kernel starch content between these NILs was observed at 14 DAP, whereas the starch content of NIL_B is significantly higher than that of NIL_A at 21 DAP (3.2%) and at maturity (2.7%) (Figure 1B). This indicates that some of the genes differentially expressed between these NILs during the period from 14 to 21 DAP might be the causal gene(s) underlying *qHS3*.

Transcriptome profile based on RNA-Seq in NILs

To obtain transcriptome profiles of maize kernels in NIL_A and NIL_B, we performed RNA-Seq on 14 and 21 DAP maize kernels of both NILs. A total of 12 libraries, including three biological replicates for each line at each stage, were prepared for RNA-Seq. A total of 190 million 50-bp single-end reads were obtained for 12 samples, with reads for each sample ranging from 10 million to 23 million (Table S1). The read length for all samples combined is >9 Gb, representing a approximately 3.7-fold coverage of the maize genome. Approximately 81% of these short reads could be uniquely mapped to the reference B73 genome (Version 5b.60) (Table S1). Deep sequencing of the maize transcriptome covered 25,201 genes in all NILs, and 98.0% (24,701/25,201) of all genes were detected in both NIL_A and NIL_B at both developmental stages (Figure 2A; Table S2). In addition, the number of genes uniquely expressed in NIL_A or NIL_B at each stage ranges from 7 to 97 (Table S2).

To evaluate the effect of the varying conditions on transcript expression profiles of NIL_A and NIL_B at 14 and 21 DAP, principal component analysis (PCA) was performed using the raw expression data (RPKM) of all identified genes in the 12 samples. This clustering reveals the relatedness of all datasets in these samples (Figure 2B). The NILs at each developmental stage clusters as nearest neighbors, indicating dramatic expression differences between the two stages. Additionally, NIL_A and NIL_B are distant at 14 DAP but cluster together at 21 DAP, suggesting that the level of differential gene expression between NILs at 14 DAP is higher than that at 21 DAP. All three biological replicates cluster as nearest neighbors from each other in most cases, indicating slight variability among the biological replicates.

Genomic variations between NILs

To deeply mine the genomic variations between NILs, polymorphic SNPs were detected using all RNA-Seq data, which identified 659 SNPs between the two NILs, with 658 SNPs distributed in the introgressed fragment on chromosome 3 and 1 SNP located on chromosome 6 (Figure 3A; Table S3), consistent with results inferred using the MaizeSNP3K array (Figure 1A). By combining data from the array and RNA-Seq, the two NILs were found to mainly differ in the 173-Mb genomic fragment ranging from 16,566,519 to 189,901,888 bp on chromosome 3 with 709 SNPs. Of these 709 SNPs, 678 are available for predicting effects on gene

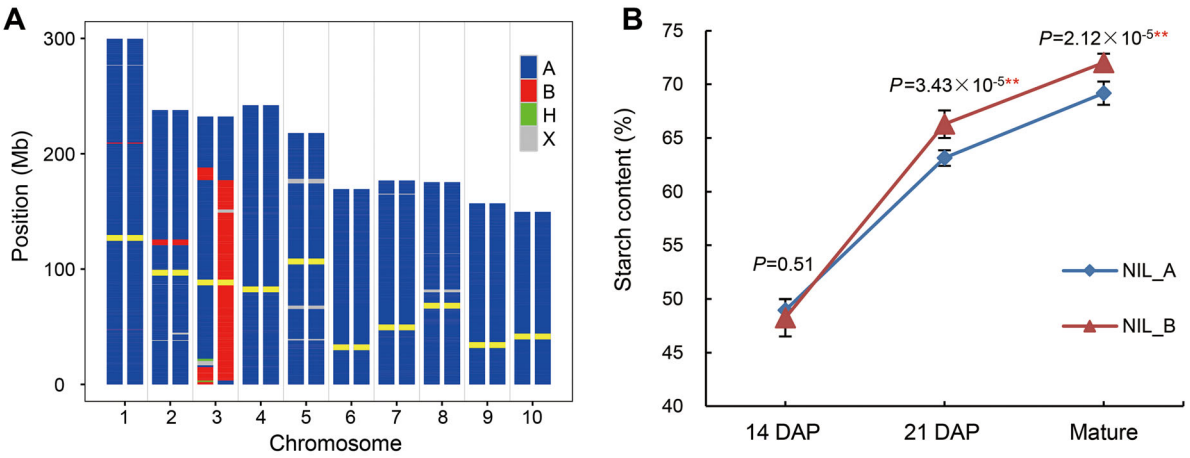

Figure 1. Characteristics of NIL_A and NIL_B
(**A**) Genomic composition of NIL_A (left) and NIL_B (right). Blue represents regions homozygous with B73; red represents homozygous regions from DHLoPro1; green represents heterozygous regions; grey represents missing data; yellow lines indicate the centromeres. (**B**) Comparison of starch accumulation in maize kernels at different developmental stages. Values represent the mean ± *SE* ($n = 15$). *P*-values were estimated based on the Student's *t*-test.

function, including 261 SNPs in the 5′ and 3′ untranslated regions or in introns, 295 synonymous and 122 nonsynonymous SNPs in the coding region (Table S3). One nonsynonymous SNP resides in an annotated gene, GRMZM2G068913 (*ZmHXK3a*), which is involved in starch metabolism. *ZmHXK3a* encodes a hexokinase, which converts glucose to glucose-6-phosphate (G6P; Figure S1), a step in starch synthesis (Tsai et al. 1970; Doehlert et al. 1988).

Transcriptomic variations between NILs

Besides the observed genomic variations between NILs, the expression patterns of genes within the mapping interval could be used to prioritize candidate causal genes as they often affect phenotypic variation via transcriptional regulation. We compared gene expression in developing kernels between NIL_B and NIL_A at 14 and 21 DAP, and differentially expressed genes (DEGs) were defined as those up- or downregulated by ≥1.5-fold ($P < 0.05$, Fisher's test). In total, 2,920 genes, evenly distributed across all chromosomes with 374 genes in the introgressed fragment interval, are differentially expressed between the NILs at 14 DAP; 148 genes are upregulated, and 226 are downregulated (Figure 3A, B; Table S4). Far fewer DEGs were detected at 21 DAP, with 59 being upregulated and 64 downregulated (Figure 3A, B; Table S4). Approximately 71% (87/123) of these DEGs fell within the introgressed fragment interval. Among all

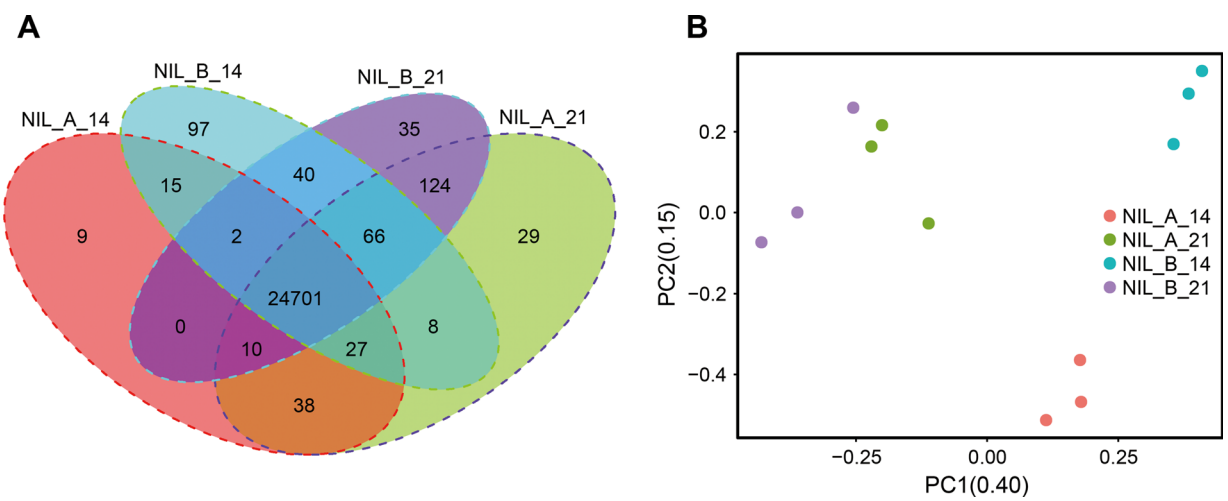

Figure 2. Transcriptomic profiles of near-isogenic lines (NILs)
(**A**) Euler diagrams of shared presence or absence of genes in NILs at two developmental stages. (**B**) Clusters of NILs at different developmental stages identified by principal component analysis of gene expression level (RPKM). Each spot represents an individual sample, with three biological replicates, indicated by the name of each line followed by the developmental stage.

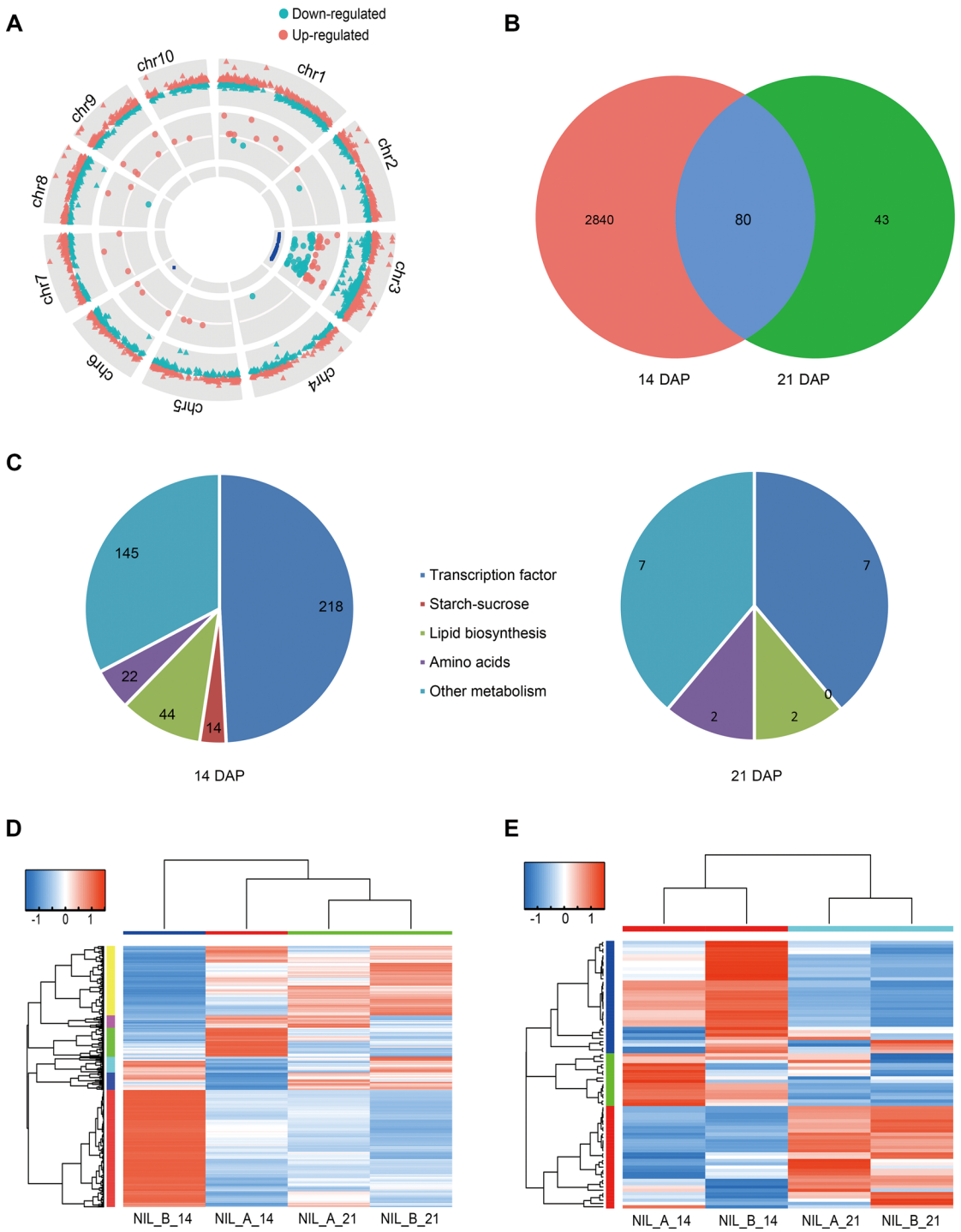

Figure 3. Comparative genomics and transcriptomics of near-isogenic lines (NILs)
(**A**) Distribution of single nucleotide polymorphisms (SNPs) between NILs, and differentially expressed genes (DEGs) between NILs at different developmental stages. The three circles from inner to outer show the distribution of SNPs identified by RNA-Seq, DEGs at 21 d after pollination (DAP), and DEGs at 14 DAP, respectively. The dark red dots above of the line represent upregulated genes, while the blue dots under the line represent downregulated genes. (**B**) Euler diagrams of DEGs between NILs at two developing stages. (**C**) Association between gene function and DEGs at two different stages. Only genes encoding transcription factors and enzymes in metabolic pathways are shown. (**D**) Self-organizing heatmap clustering DEGs by their expression abundance (represented as normalized RPKM with Z-score). Red represents high expression level, and blue represents low abundance. (**E**) Self-organizing heatmap clustering 81 genes known or predicted in starch metabolic pathways by their expression abundance (represented as normalized RPKM with Z-scores). Red represents high abundance, and blue represents low abundance.

the DEGs, 80 are common at both 14 and 21 DAP, including 77 DEGs which have expression level differences in the same direction at both developmental stages. Heat map clustering also yielded an overview of the expression of the 2,963 DEGs with six distinct patterns in the two NILs at both stages (Figures 3D, S2; Table S4). In cluster I, the expression levels of DEGs between the two NILs at 14 DAP (e.g., GRMZM2G325920) decrease sharply in NIL_B whereas there is little change in NIL_A during development. A similar contrast was observed for clusters II and III (e.g., GRMZM2G127798 and GRMZM2G479110); there are no significant differences for the DEGs between the two NILs at 14 DAP in NIL_B or NIL_A at 14 DAP during the developing stages. For the remaining three clusters, the expression levels of the DEGs between the two NILs differ significantly at both stages, but the direction and degree of difference vary among clusters. Compared with NIL_A, the genes in NIL_B are enriched in cluster IV but reduced in cluster VI at both stages, whereas they are upregulated at 14 DAP but downregulated at 21 DAP in cluster V.

To determine whether the 2,963 DEGs at 14 and 21 DAP are associated with starch biosynthesis and accumulation, their putative functions were annotated using MapMan software (Thimm et al. 2004). Gene set enrichment analysis revealed significantly enriched terms (false-discovery rate < 0.05) in the categories of DNA, RNA, protein, cell and development, responses to stress, metabolic process and etc. (Table S4). Over half of the DEGs at 14 DAP in assigned categories are enriched in the terms related to the activities of DNA, RNA, protein, cell and development, suggesting a possible role in the development and growth in maize kernel at 14 DAP. Subsequent analysis focused on genes related to transcription factors and metabolic processes (Figure 3C) because they would more likely be associated with starch biosynthesis and accumulation. There are 218 transcription factors with differential expression at 14 DAP, including 25 within the introgressed fragment interval, which correlate well with the large number of DEGs at 14 DAP. In contrast, there is a sharp decline in the number of differentially expressed transcription factors at 21 DAP, and only five of them fall within the introgressed fragment interval. Similar patterns were also observed for other metabolic processes, but the magnitude of the changes in the metabolic process DEGs is much smaller than in the transcription factors category.

Validation of the identified DEGs

To confirm the regulation of the identified DEGs, we compared the expression level of 20,342 genes in NIL_A with publically available transcriptomic data for B73 kernels (Chen et al. 2014). Nearly 70% of the genes share the same expression patterns, demonstrating the reliability of the RNA-Seq data in our study. We also selected 10 genes involved in starch metabolism and subjected them to quantitative real-time (qRT)-PCR using the two NILs at two stages (Figure 4). A close correspondence was observed between expression levels detected by RNA-Seq and by qRT-PCR ($r = 0.93$) (Figure S3), thus further validating the RNA-Seq analysis.

Expression pattern of genes in starch metabolism

To gain insight into the expression pattern of genes involved in starch metabolism in the two NILs, we identified 128 genes related to this pathway in maize through manual curation of the MapMan, MaizeCyc, and KEGG databases (Kanehisa and Goto 2000; Usadel et al. 2009). A total of 81 genes related to starch metabolism were identified in the expression data in kernels at 14 and 21 DAP, including 42 genes involved in starch biosynthesis and 39 involved in starch degradation (Table S5). These 81 genes were grouped into three clusters (Figure 3E). Type I, with most of the genes related to starch biosynthesis, has higher expression levels at 14 DAP than at 21 DAP in both NILs. In contrast, type III, which contains most of the genes related to starch degradation, has lower expression levels at 14 DAP than at 21 DAP. These results were consistent with higher levels of starch synthesis at early stages, with starch degradation being activated at later stages. Type II genes, which have similar expression levels at both stages, contain both biosynthesis and degradation genes.

Among 81 available genes in starch metabolism, seven DEGs between both NILs were detected at 14 DAP and none at 21 DAP, including *ZmHXK3a* (GRMZM2G068913) in the introgressed fragment interval, *AGPLLZM* (GRMZM2G391936) on chromosome 1, *SSV* (GRMZM2G130043) on chromosome 4, *ZmGPE5* (GRMZM2G011662) on chromosome 6, *GBSSII* (GRMZM2G008263) on chromosome 7, *ZmHXK3b* (GRMZM2G467069) on chromosome 8, and *Du1* (GRMZM2G141399) on chromosome 10 (Figure 5). Four DEGs, *ZmHXK3a*, *SSV*, *ZmHXK3b* and *Du1*, are upregulated in NIL_B while three DEGs, *AGPLLZM*, *ZmGPE5* and *GBSSII*, are downregulated, which are mostly consistent with the roles of the encoded enzymes in starch metabolism. Upregulation of *ZmHXK3a* and *ZmHXK3b* would result in more G6P, thereby providing more substrate for the synthesis of ADP-glucose and thus accelerating starch metabolic processes. As expected, the expression of *Du1*, encoding starch synthase, is upregulated. A newly identified gene, *SSV*, also encoding a starch synthase, has the same expression pattern as *Du1* likely owing to their similar functions in starch synthesis. In contrast, *GBSSII*, encoding a granule-bound starch synthase, has the opposite expression pattern of *Du1* and *SSV*. This may be because *Du1* and *SSV* catalyze 1-4-α-glucose to produce amylopectin whereas *GBSSII* catalyzes 1-4-α-glucose to produce amylose. These results confirmed the effect of *qHS3* on starch content in maize kernel. The expression of *AGPLLZM*, encoding the AGPase large subunit, is down-regulated, which might be due to negative feedback. *ZmGPE5*, encoding a glucose-6-phosphate 1-epimerase, is also downregulated which might have little effect on the starch content because of its role in balancing two intermediate products in the starch biosynthesis.

DISCUSSION

The molecular mechanisms of starch biosynthesis and accumulation in maize kernels are central for improving grain yield and quality. Previous studies have attempted to define a clear framework for starch metabolic pathways in maize and other cereals (Jeon et al. 2010; Walley et al. 2013). Additional knowledge about maize starch metabolism, and particularly its regulation, could have a substantial impact on maize grain quality. In this study, the genomic and transcriptomic analyses of developing maize kernels at 14 and 21 DAP from a pair of

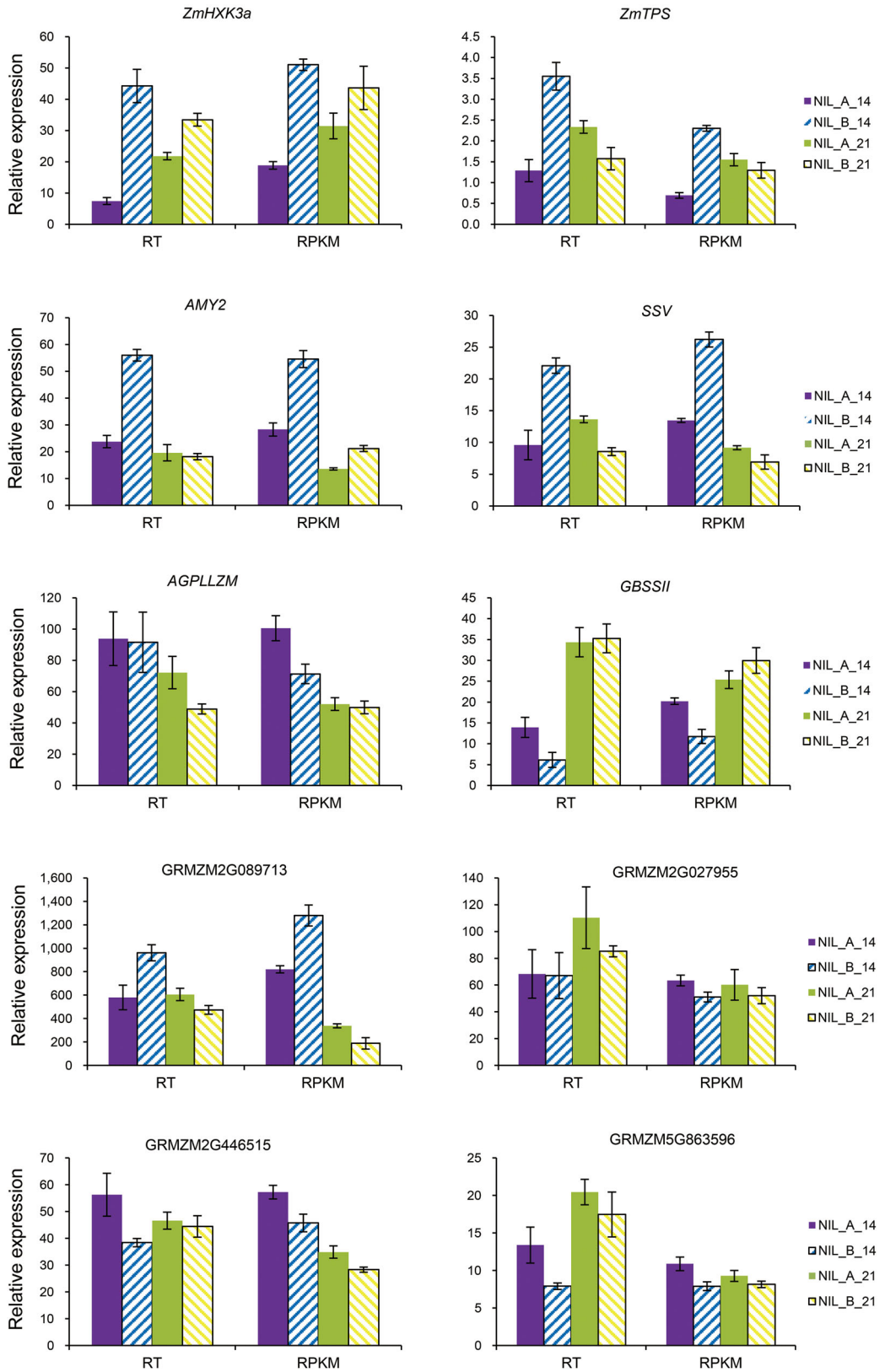

Figure 4. Relative expression ratio between near-isogenic lines (NILs) of 10 genes involved in the starch metabolic pathway in kernels at different developmental stages (14 and 21 d after pollination)

Expression levels were normalized to ACTIN. For each plot, the left columns represent the results in qPCR, while the right columns represent RPKM in RNA-Seq. Error bars represent *SE* ($n = 3$).

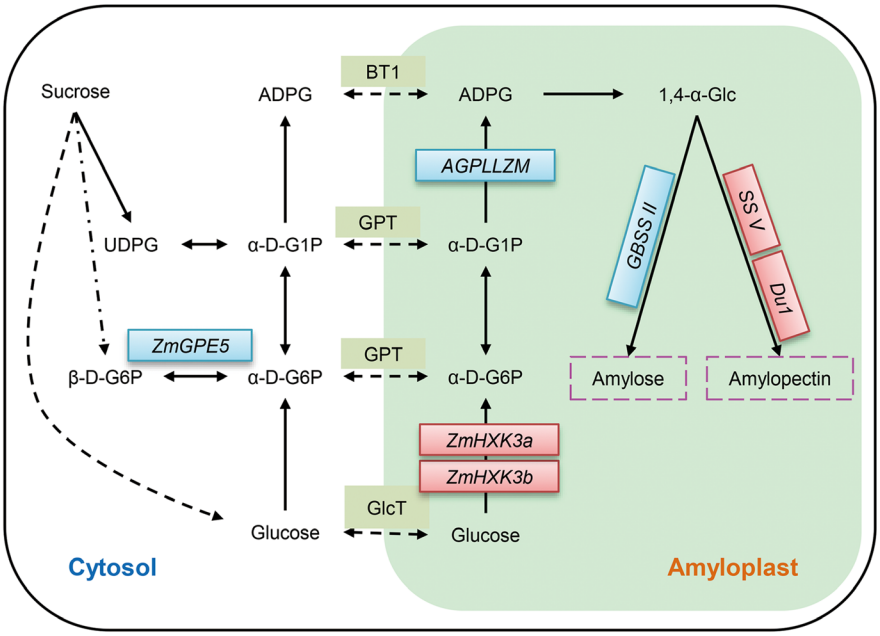

Figure 5. Overview of the coexpression network caused by *qHS3* in maize kernel
Seven differentially expressed genes (DEGs) are highlighted in the starch biosynthesis pathway Red represents upregulated genes, and blue represents downregulated genes. The cytosol and amyloplast compartments are indicated. α-D-G1P, α-D glucose-1-phosphate; α-D-G6P, α-D glucose-6-phosphate; β-D-G6P, β-D glucose-6-phosphate; ADPG, ADP-glucose; BT1, ADP-glucose transporter; GPT, glucose 1/6-phosphate transporter; GlcT, glucose transporter; UDPG, UDP-glucose; UGPase, UDP-glucose pyrophosphorylase.

NILs with different starch content indicate that a number of genes might be responsible for the variable starch contents caused by *qHS3* on chromosome 3. The altered regulation of the candidate genes in *qHS3* in NIL_B at 14 DAP lead to higher activities of several key enzymes in the starch metabolic pathway, which likely resulted in increased starch content.

Potential causal genes for *qHS3*

Whole-transcriptome sequencing approaches, such as RNA-Seq, as well as the whole-genome sequencing approaches have been a popular method for identifying the causal genetic factors controlling phenotypic variation since Ng et al. (2010) reported the first successful application of exome sequencing to discover a causative mutation. Because RNA-Seq provides information on both the genomic and transcriptomic variation of a mutant or QTL, it is an efficient strategy for mapping genes and understanding their transcriptional regulation. Recently, Barrero et al. (2015) identified two adjacent genes, *PM19-A1* and *A2*, as candidates for a major dormancy QTL by transcriptomic analysis of wheat NILs. This underscores the possibility of identifying candidate genes underlying given QTL in plants by RNA-Seq. In this study, 76 genes with nonsynonymous SNPs and 384 DEGs in the introgressed fragment, with 22 genes having both nonsynonymous mutations and differential expression (Tables S3, S4), might be responsible for *qHS3*. According to the functional annotation of these genes, only GRMZM2G068913 (*ZmHXK3a*), encodes a hexokinase involved in starch metabolism, making it a potential causative gene for elevated starch content in NIL_B. A G-to-A transition occurs in the exon 9 of *ZmHXK3a*, which converts glutamic acid to lysine at amino acid residue 471. In addition, *ZmHXK3a* is upregulated in NIL_B at both 14 and 21 DAP (Figure S3; Table S4), which also supports that *ZmHXK3a* might be one of the important candidate genes for *qHS3*.

ZmHXK3a encodes a key enzyme, converting glucose to G6P, and therefore may play important roles in adjusting the rate of starch synthesis, due to G6P's role as an important intermediate product in starch metabolism. Intensive investigations of hexokinases in numerous plant species have revealed their distinct roles in plant metabolism, including controlling starch content (Veramendi et al. 1999; Giese et al. 2005). In maize, nine *ZmHXK* gene family members have been identified (Zhang et al. 2014). Two of these genes, *ZmHXK3a* and *ZmHXK3b*, are homologous to *OsHXK3* in rice (Zhang et al. 2014). They are both upregulated in NIL_B kernels at 14 DAP, suggesting that both of these hexokinase genes could exhibit the same function in controlling starch content. The fact that *ZmHXK3b* has the same expression pattern as *ZmHXK3a* but does not have a QTL associated with it indicates that the regulation of *ZmHXK3b* is likely a secondary effect of a gene within the *qHS3* interval. To validate whether *ZmHXK3a* controls *qHS3*, one direct way is to overexpress or knockdown *ZmHXK3a* to see whether the starch content in maize kernel will change. In addition, the mechanism of *ZmHXK3a* regulating starch variation could occur at the post-transcriptional or protein level. Thus, analyzing the enzyme activity of different *ZmHXK3a* alleles may also validate whether *ZmHXK3a* controls *qHS3*.

The maximum difference in kernel starch content between NILs occurred at 21 DAP, whereas no significant difference was seen at 14 DAP, suggesting that the transcriptional regulation of starch metabolism likely impacts earlier stages. Thus, the causal genes for *qHS3* could include DEGs involved in starch metabolism in the *qHS3* interval or genes in that interval regulating DEGs in starch metabolism outside the QTL interval at 14 DAP. Therefore, 25 transcription factors with differential expression at 14 DAP are the second set of genes that may be responsible for *qHS3*. To determine whether these genes are the real causal genes for *qHS3*, fine mapping of *qHS3* is necessary to narrow down the QTL interval.

Transcriptional effect of the candidate genes for *qHS3*

Transcriptional regulation is one of the molecular mechanisms by which genes affect phenotypic variation. Altering the regulation of some candidate genes in the *qHS3* region could therefore lead to transcriptional responses in related genes. The large fragment introgressed in NIL_B caused a large number of genes differentially expressed, which implies a complex regulatory framework involving the candidate genes in *qHS3* region. Five more DEGs, *AGPLLZM*, *SSV*, *GBSSII Du1* and *ZmGPE5* were identified besides *ZmHXK3a* and *ZmHXK3b*. The expression pattern of all DEGs caused by the candidate genes for *qHS3* is consistent with their activities in starch metabolism besides *AGPLLZM*, which is one large subunit of AGPase.

One of the rate-limiting reactions in starch metabolism is the pyrophosphorylation of glucose-1-phosphate by AGPase, which forms ADP-glucose, the precursor of starch synthesis. AGPase is a heterotetramer composed of two large subunits and two small subunits, each of which is encoded by distinct genes (Hannah and James 2008). In maize for example, *Shrunken2* (*Sh2*), *AGPLLZM*, and *AGPLEMZM* encode the large subunits, and *Brittle2a* (*Bt2a*), *Brittle2b* (*Bt2b*), *AGPSEMZM*, *AGPSLZM*, and *AGPL3* encode the small subunits (Huang et al. 2014). The mutant characterization of *sh2*, *bt2*, *agpsemzm* and *agpllzm* in maize kernels provides evidence that downregulation of these three genes reduces AGPase activity and thus decreases starch content (Giroux and Hannah 1994; Huang et al. 2014). However, all seven available genes encoding subunits of AGPase are downregulated in NIL_B at 14 DAP except for *AGPSLZM*, although *AGPLLZM* was the only statistically significant DEG among these genes (Table S5). This makes sense, given that *AGPLLZM* is the plastidial form of AGPase, and its downregulation has no effect on the total endosperm AGPase activity since around 95% of the total AGPase in maize endosperm is extraplastidial (Denyer et al. 1996; Huang et al. 2014).

A new gene identified in starch metabolism

Besides coexpression network in a specific metabolic pathway, RNA-Seq also provides an opportunity to mine new genes controlling a target trait and explain their functions. For example, Yang et al. (2013b) identified five genes that participate in maize photoperiod pathway by RNA-sequencing the leaf tissue at the three-leaf stage from a pair of NILs for photoperiod sensitivity. The total number of genes encoding starch synthase is unknown in maize, although eight genes, namely *Du1*, *SS1*, *Su2*, *SS2b-2*, *SS2c*, *SS3b-1*, *SS3b-2*, and *SS4*, were cloned using mutants in maize or via comparison with the rice genome (Gao et al. 1998; Harn et al. 1998; Knight et al. 1998; Yan et al. 2009a,b). Here, we detected a new starch synthase gene, *SSV*, upregulated in NIL_B as well as the known starch synthase gene, *Du1*. Their similar expression patterns indicate that they may have similar functions in starch biosynthesis.

MATERIALS AND METHODS

Plant materials

The BC_5F_2 NIL parir maize (*Zea mays* L.) was grown with three biological replicates in the winter nursery of China Agricultural University in 2011 in SanYa, Hainan province, China. The bulked leaves of each NIL were collected from at least six individuals for DNA extraction. All plants were self-pollinated, and four immature ears were harvested at 14 and 21 DAP for each NIL in each replicate. Four kernels from each ear of one NIL within a replicate were randomly chosen and stored at –80 °C for RNA extraction. The remaining kernels of each ear were dried for measurement of starch content. Mature ears of NILs were also collected for analysis of starch.

Measurement of starch content

In each biological replicate, a total of 15 ears were used to measure starch content in both NILs. Starch content in maize kernels was determined with a fermentable carbohydrate assay (Zhou and Bao 2012). Fifteen immature or mature kernels from the middle section of each ear were ground to fine powder by Geno 2010 Grinder (Spex, Metuchen, NJ, USA). Around 120 mg powder was placed into a 2 mL snap cap tube. An α-amylase solution was added to each tube, then the tubes were placed in a boiling water bath for 15 min. Tubes were then incubated at 50 °C for 1.5 h, followed by addition of Glucoamylase. Finally, 3 mg yeast was added to ferment glucose into ethanol and carbon dioxide. Starch content was calculated as the weight loss due to fermentation (CO_2) and heat (ethanol). All ears were analyzed with fermentable carbohydrate assay for twice, and the average was taken as the starch content.

DNA extraction and genotyping

DNA was extracted from leaf tissue of both NILs using the modified CTAB method (Murray and Thompson 1980) and genotyped with the MaizeSNP3K array containing 3,072 SNPs, a subset of the Illumina MaizeSNP50 Beadchip (Ganal et al. 2011). SNP genotyping was performed on the Illumina GoldenGate SNP genotyping platform at the National Maize Improvement Center of China, China Agricultural University. The quality of each SNP was assessed manually, and SNPs with poor quality were excluded from further analysis.

RNA extraction and sequencing

Total RNA was isolated from frozen maize kernels with the Qiagen RNeasy kit (Qiagen, Valencia, CA, USA). RNA-Seq libraries with 150- to 200-bp insert size were prepared using the Illumina TruSeq mRNA-Seq kit and sequenced on the Illumina HiSeq 2000 system with TruSeq SBS v3 reagents based on the protocol from Illumina (San Diego, CA, USA).

RNA-Seq mapping, SNP calling and annotation

After removing reads containing sequencing adapters and poly-A as well as reads of low sequencing quality (missing rate $\geq$ 0.001) through perl scripts, TopHat v2.0.9 (Trapnell et al. 2012) was used to map the single-end clean reads against the B73 reference genome sequence (B73 AGPv2; http://www.maizesequence.org). Only reads that mapped uniquely to the genome were retained for further analysis.

The uniquely mapped reads were sorted depending on their alignment position on each chromosome and then converted to SAM format using Picard tools and Samtools (Li et al. 2009). Unpaired SNP calling was performed using the Genome Analysis Toolkit pipeline (DePristo et al. 2011). SNPs detected as the same in all three biological replicates and kept consistent in the same NIL at two developmental stages were used for subsequent analysis. SNPs were categorized as intergenic, exonic, or intronic based on their position. The genetic codes in which exonic SNPs were located were extracted to classify them as synonymous or nonsynonymous.

Quantification of gene expression

The number of the uniquely mapped reads in each gene was counted using HTSeq v0.5.4p3 (Mortazavi et al. 2008), and the RPKM of each gene was calculated based on the length of the gene and reads mapped to it (Mortazavi et al. 2008). Only genes with RPKM showing consistent trends ($>$0 or $=$0) within three biological replicates simultaneously were included for further analysis.

Identification of DEGs

To identify differentially expressed genes (DEGs), a model based on the negative binomial distribution was performed using DESeq R package (1.10.1) with default settings (Anders and Huber 2012). The false-discovery rate was controlled by the approach of Benjamini and Hochberg (1995). Gene expression levels were considered to differ significantly if (i) the adjusted *P*-value was $<$0.05 in a *t*-test and (ii) the fold-change was $>$1.5 between two samples (Anders and Huber 2012).

Principal component analysis and hierarchical clustering

Principal component analysis was performed using a custom R script to analyze the raw expression values (i.e., RPKM) for all identified genes in 12 samples. For the DEGs and the key genes in the starch metabolic pathway, hierarchical clustering was performed using Euclidean distance metric in the R script. For a given gene, Z-scores of the average RPKM in three replicates were also calculated.

Functional gene classification and GO enrichment

The software MapMan (http://mapman.gabipd.org) was used to annotate genes. A total of 35 Map Bins were classified in MapMan, including 40,843 genes in the maize genome to facilitate the classification of DEGs. Gene ontology (GO) enrichment for DEGs was also analyzed with MapMan. The number of genes in each Bin was counted. Fisher's exact test was used to investigate the GO enrichment category. *P*-values were adjusted for multiple testing based on the false-discovery rate with the BH method (Benjamini and Hochberg 1995). Bins for which the false-discovery rate was $<$0.05 were considered to be enriched.

qRT-PCR

First-strand cDNA was synthesized using the ProtoScriptc DNA Synthesis kit (New England Biolabs, UK). Quantitative RT-PCR was carried out in triplicate for each sample using the SYBRGreen I kit (Takara Biotechnology, China) on a 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Maize ACTIN was used as a control for normalization between samples. Relative quantification of genes was calculated using the comparative threshold cycle method (Livak and Schmittgen 2001). Supplemental Table S6 lists primers used for qRT-PCR.

ACKNOWLEDGEMENTS

We appreciate George Singletary, Lan Zhou, Junjie Fu, Thomas Moloshok, Charlotte Martin and Brandon Van Allen for their technical support. This study was supported by the National Natural Science Foundation of China (31421005), International Cooperation in Science and Technology Project in China (2014DFG31690) and DuPont Pioneer.

AUTHOR CONTRIBUTION

X.Y. performed the experiments, data analysis and wrote the manuscript; T.S., W.M. and L.L. performed the data analysis; W.T. performed the field experiments; B.M. and Z.G. performed the RNA-Seq; L.J. and L.B. co-designed the study; Y.X. designed the study and wrote the manuscript.

REFERENCES

Anders S, Huber W (2012) *Differential Expression of RNA-Seq Data at the Gene Level–the DESeq Package*. EMBL, Heidelberg, Germany

Barrero JM, Cavanagh C, Verbyla KL, Tibbits JF, Verbyla AP, Huang BE, Rosewarne GM, Stephen S, Wang P, Whan A, Rigault P, Hayden MJ, Gubler F (2015) Transcriptomic analysis of wheat near-isogenic lines identifies PM19-A1 and A2 as candidates for a major dormancy QTL. **Genome Biol** 16: 93

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. **J Roy Stat Soc B Met** 57: 289–300

Chen J, Zeng B, Zhang M, Xie S, Wang G, Hauck A, Lai J (2014) Dynamic transcriptome landscape of maize embryo and endosperm development. **Plant Physiol** 166: 252–264

Denyer K, Dunlap F, Thorbjørnsen T, Keeling PL, Smith AM (1996) The major form of ADP-glucose pyrophosphorylase in maize endosperm is extra-plastidial. **Plant Physiol** 112: 779–785

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. **Nat Genet** 43: 491–498

Doehlert DC, Kuo TM, Felker FC (1988) Enzymes of sucrose and hexose metabolism in developing kernels of two inbreds of maize. **Plant Physiol** 86: 1013–1019

Ellis RP, Cochrane MP, Dale MFB, Duffus CM, Lynn A, Morrison IM, Prentice RDM, Swanston JS, Tiller SA (1998) Starch production and industrial use. **J Sci Food Agr** 77: 289–311

Fu FF, Xue HW (2010) Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a

novel rice starch biosynthesis regulator. **Plant Physiol** 154: 927–938

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J (2011) A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. **PLoS ONE** 6: e28334

Gao M, Wanat J, Stinard PS, James MG, Myers AM (1998) Characterization of *dull1*, a maize gene coding for a novel starch synthase. **Plant Cell** 10: 399–412

Giese JO, Herbers K, Hoffmann M, Klösgen RB, Sonnewald U (2005) Isolation and functional characterization of a novel plastidic hexokinase from *Nicotiana tabacum*. **FEBS Lett** 579: 827–831

Giroux M, Hannah L (1994) ADP-glucose pyrophosphorylase in shrunken-2 and brittle-2 mutants of maize. **Mol Gen Genet** 243: 400–408

Goldman I, Rocheford T, Dudley J (1993) Quantitative trait loci influencing protein and starch concentration in the Illinois long term selection maize strains. **Theor Appl Genet** 87: 217–224

Hannah LC, James M (2008) The complexities of starch biosynthesis in cereal endosperms. **Curr Opin Biotech** 19: 160–165

Harn C, Knight M, Ramakrishnan A, Guan H, Keeling PL, Wasserman BP (1998) Isolation and characterization of the zSSIIa and zSSIIb starch synthase cDNA clones from maize endosperm. **Plant Mol Biol** 37: 639–649

Hennen-Bierwagen TA, Myers AM, Becraft P (2013) Genomic specification of starch biosynthesis in maize endosperm. **Seed Genomics**: 123–137

Huang B, Hennen-Bierwagen TA, Myers AM (2014) Functions of multiple genes encoding ADP-glucose pyrophosphorylase subunits in maize endosperm, embryo, and leaf. **Plant Physiol** 164: 596–611

Jeon JS, Ryoo N, Hahn TR, Walia H, Nakamura Y (2010) Starch biosynthesis in cereal endosperm. **Plant Physiol Biochem** 48: 383–392

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. **Nucleic Acids Res** 28: 27–30

Knight ME, Harn C, Lilley CE, Guan H, Singletary GW, Mu-Forster C, Wasserman BP, Keeling PL (1998) Molecular cloning of starch synthase I from maize (W64) endosperm and expression in Escherichia coli. **Plant J** 14: 613–622

Lai J, Dey N, Kim C-S, Bharti AK, Rudd S, Mayer KF, Larkins BA, Becraft P, Messing J (2004) Characterization of the maize endosperm transcriptome and its comparison to the rice genome. **Genome Res** 14: 1932–1937

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. **Bioinformatics** 25: 2078–2079

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. **Methods** 25: 402–408

Lu X, Chen D, Shu D, Zhang Z, Wang W, Klukas C, Chen L-l, Fan Y, Chen M, Zhang C (2013) The differential transcription network between embryo and endosperm in the early developing maize seed. **Plant Physiol** 162: 440–455

Méchin V, Thévenot C, Le Guilloux M, Prioul JL, Damerval C (2007) Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. **Plant Physiol** 143: 1203–1219

Martin C, Smith AM (1995) Starch biosynthesis. **Plant Cell** 7: 971

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nat Methods** 5: 621–628

Murray M, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. **Nucleic Acids Res** 8: 4321–4326

Myers AM, Morell MK, James MG, Ball SG (2000) Recent progress toward understanding biosynthesis of the amylopectin crystal. **Plant Physiol** 122: 989–998

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA (2010) Exome sequencing identifies the cause of a mendelian disorder. **Nat Genet** 42: 30–35

Prioul JL, Méchin V, Lessard P, Thévenot C, Grimmer M, Chateau-Joubert S, Coates S, Hartings H, Kloiber-Maitz M, Murigneux A (2008) A joint transcriptomic, proteomic and metabolic analysis of maize endosperm development and starch filling. **Plant Biotechnol J** 6: 855–869

Séne M, Causse M, Damerval C, Thévenot C, Prioul JL (2000) Quantitative trait loci affecting amylose, amylopectin and starch content in maize recombinant inbred lines. **Plant Physiol Biochem** 38: 459–472

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) Mapman: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. **Plant J** 37: 914–939

Tsai C, Salamini F, Nelson O (1970) Enzymes of carbohydrate metabolism in the developing endosperm of maize. **Plant Physiol** 46: 299–306

Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M (2009) A guide to using MapMan to visualize and compare Omics data in plants: A case study in the crop species, Maize. **Plant Cell Environ** 32: 1211–1229

Veramendi J, Roessner U, Renz A, Willmitzer L, Trethewey RN (1999) Antisense repression of hexokinase 1 leads to an overaccumulation of starch in leaves of transgenic potato plants but not to significant changes in tuber carbohydrate metabolism. **Plant Physiol** 121: 123–134

Walley JW, Shen Z, Sartor R, Wu KJ, Osborn J, Smith LG, Briggs SP (2013) Reconstruction of protein networks from an atlas of maize seed proteotypes. **Proc Natl Acad Sci USA** 110: E4808–E4817

Wassom JJ, Wong JC, Martinez E, King JJ, DeBaene J, Hotchkiss JR, Mikkilineni V, Bohn MO, Rocheford TR (2008) QTL Associated with maize kernel oil, protein, and starch concentrations; kernel mass; and grain yield in Illinois high oil × B73 backcross-derived lines. **Crop Sci** 48: 243

Yan HB, Pan XX, Jiang HW, Wu GJ (2009b) Comparison of the starch synthesis genes between maize and rice: Copies, chromosome location and expression divergence. **Theor Appl Genet** 119: 815–825

Yan H, Jiang H, Pan X, Li M, Chen Y, Wu G (2009a) The gene encoding starch synthase IIc exists in maize and wheat. **Plant Sci** 176: 51–57

Yang G, Dong Y, Li Y, Wang Q, Shi Q, Zhou Q (2013a) Verification of QTL for grain starch content and its genetic correlation with oil content using two connected RIL populations in high-oil maize. **PloS ONE** 8: e53770

Yang Q, Li Z, Li WQ, Ku LX, Wang C, Ye JR, Li K, Yang N, Li YP, Zhong T (2013b) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. **Proc Natl Acad Sci USA** 110: 16969–16974

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K (2008) Comprehensive flavonol profiling and transcriptome

coexpression analysis leading to decoding gene–metabolite correlations in *Arabidopsis*. **Plant Cell** 20: 2160–2176

Zeeman SC, Kossmann J, Smith AM (2010) Starch: Its metabolism, evolution, and biotechnological modification in plants. **Annu Rev Plant Biol** 61: 209–234

Zhang J, Lu X, Song X, Yan J, Song T, Dai J, Rocheford T, Li J (2008) Mapping quantitative trait loci for oil, starch, and protein concentrations in grain with high-oil maize by SSR markers. **Euphytica** 162: 335–344

Zhang Z, Zhang J, Chen Y, Li R, Wang H, Ding L, Wei J (2014) Isolation, structural analysis, and expression characteristics of the maize (*Zea mays* L.) hexokinase gene family. **Mol Biol Rep** 41: 6157–6166

Zhou L, Bao X (2012) High throughput method for measuring total fermentables in small amount of plant part. **U.S. Patent** 8,329,426 [P]. 2012-12-11

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Figure S1. A simplified starch metabolic pathway in maize kernel, modified from pathways defined for other cereals
The cytosol and amyloplast compartments are indicated. α-D-G1P, α-D glucose-1-phosphate; α-D-G6P, α-D glucose-6-phosphate; β-D-G6P, β-D glucose-6-phosphate; ADPG, ADP-glucose; AGPase, ADP-glucose pyrophosphorylase; BT1, ADP-glucose transporter; GBSS, Granule-bound starch synthase; GlcT, glucose transporter; GPE, glucose-6-phosphate 1-epimerase; GPT, glucose 1/6-phosphate transporter; HXK, hexokinase; PGMase, phosphoglucomutase; SBE, starch-branching enzyme; SS, starch synthase; SuSy, sucrose synthase; UDPG, UDP-glucose; UGPase, UDP-glucose pyrophosphorylase.

Figure S2. Expression patterns of differentially expressed genes (DEGs) between NIL_A and NIL_B at two stages

Figure S3. Plots of the correlation between expression level estimated by RNA-Seq and RT-PCR
The x axis represents relative expression values from RT-PCR. The y axis represents the RPKM values from RNA-Seq. Both values are normalized with the logarithm to base 10. The r-value is a Pearson correlation coefficient.

Table S1. Summary of RNA-Seq yields and alignments for 12 samples

Table S2. List of genes expressed in 12 samples

Table S3. Predicted effects of SNPs polymorphic between NILs identified by RNA-Seq and Array

Table S4. List of DEGs between NILs

Table S5. List of all genes involved in starch metabolism

Table S6. List of PCR primers used in this study